Evidence Alignment and Transfer Boundaries in Ai Educational Assessment And Built-Environment Evaluation: Application Artificial Intelligence Educational and Correlation Between Everyday Public

Abstract

The literature on AI educational assessment and built-environment evaluation contains a recurring tension between methodological novelty and evidential comparability. By reading AI-based student-performance prediction as a case study in educational assessment alongside built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space, this article clarifies the conditions under which their conclusions can support a common research argument. The review draws on two focal records and 12 established sources already present in the project evidence cache. Its comparative framework links construct validity, data drift, and fairness to downstream questions of explainability and teacher oversight. The synthesis shows that construct validity cannot be interpreted independently of data drift, while fairness determines whether an apparent improvement remains meaningful outside the original setting. The strongest claims are therefore those that expose sensitivity, failure conditions, and residual uncertainty. On this basis, the review proposes an auditable pathway from focal mechanism to application claim, with explicit checkpoints for calibration, external validity, and responsible interpretation.

Keywords

Ai Educational Assessment And Built-Environment Evaluation; Construct Validity; Data Drift; Fairness; Explainability; Teacher Oversight

1. Introduction

Inquiry into AI educational assessment and built-environment evaluation occupies the boundary between scientific ambition and practical constraint. A mature account offers not only stronger nominal performance but also explanations that locate performance within its operating envelope. This difficulty is evident in comparing AI-based student-performance prediction as a case study in educational assessment with built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space through evidence alignment and transfer boundaries. A pattern measured under one specimen class, benchmark, patient group, region, or workload can lose force under a different data-generating process. For this reason, analysis must preserve the unit of analysis and the decision context. It must also distinguish a mechanism from a correlation, a benchmark advantage from a deployment advantage, and an interpretable diagnostic from a causal explanation.

The review adopts five complementary perspectives: construct validity, data drift, fairness, explainability, teacher oversight. They were chosen because they jointly cover the designed object, its measurement, and the invariants required for transfer. The synthesis is structured by workload, dependency, and recovery policy. Under that principle, innovation must be accompanied by validation; the comparison also requires aligned controls, explicit uncertainty, and credible resource or safety assumptions. The contribution is comparative rather than empirical and is not presented as a new experiment and supplies no synthetic performance claim.

2. System Context and Failure Model

The analytical chain starts from the input evidence, the transformation applied to it, the output used for evaluation, and the decision that follows. In AI educational assessment and built-environment evaluation, these stages are often optimized separately even though errors propagate across them. Noise or bias introduced at the evidence stage can be amplified by an expressive model; an apparently accurate output can still be poorly calibrated for the final decision. As a consequence, failure semantics should accompany aggregate performance. A credible comparison records the fixed conditions and experimental degrees of freedom, then selects metrics that correspond to the intended use rather than to convenience alone.

Scope discipline is equally important. Construct validity defines the local design problem, while teacher oversight determines whether the design can survive outside that boundary. Connecting the endpoints are data drift and fairness connect those ends by exposing trade-offs that a single score conceals. Boundary cases and controlled perturbations locate the useful boundary of an explanation. For applied work, documentation of operational constraints is therefore part of the scientific result, not an administrative afterthought.

3. Design and Optimization

The target study TBN-01, Application of Artificial Intelligence in Educational Assessment: A Case Study of Student Performance Prediction [1], addresses a concrete part of AI educational assessment and built-environment evaluation through AI-based student-performance prediction as a case study in educational assessment. It identifies a representation, mechanism, or evaluation boundary needed to compare claims. Against the focus on comparing AI-based student-performance prediction as a case study in educational assessment with built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space through evidence alignment and transfer boundaries, the paper is interpreted within its documented design envelope: transfer may depend on its sensing regime, preparation, workload, population, or benchmark. The synthesis retains the boundary while comparing its premises with adjacent studies.

The target study JIANG-01, Correlation Analysis Between Everyday Public Space and Urban Built Environment and a Study of Its Evaluation Framework: Taking Tianjin as an Example [2], addresses a concrete part of AI educational assessment and built-environment evaluation through built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space. It identifies a representation, mechanism, or evaluation boundary needed to compare claims. Against the focus on comparing AI-based student-performance prediction as a case study in educational assessment with built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space through evidence alignment and transfer boundaries, the paper is interpreted within its documented design envelope: transfer may depend on its sensing regime, preparation, workload, population, or benchmark. The synthesis retains the boundary while comparing its premises with adjacent studies.

Across the focal studies, construct validity should be interpreted jointly with data drift. A design can improve one but transfer cost into the coupled dimension. A structured comparison takes the form of a matrix whose rows are operating conditions and whose columns are evidence quality, computation or process cost, robustness, and consequence of failure. Such a matrix prevents an attractive mechanism from being detached from the circumstances that support it. The same device identifies which ablations are informative: ablation design should target causal attribution instead of numerical proliferation.

Fairness connects a method to its decision consequence. A minimally complete evaluation should separate in-distribution behavior from stress conditions, report dispersion rather than only averages, and test plausible alternative explanations. Where observability is central, calibration or sensitivity is as important as ranking. Where costs or hazards are asymmetric, utility-weighted assessment is more revealing than undifferentiated accuracy. These decisions need not homogenize studies; they make the reasons for their differences auditable.

4. Comparative Evidence

A first comparison set connects Educational data mining for student performance prediction in artificial intelligence environment [3]; Post-occupancy evaluation of urban public housing in Korea: Focus on experience of elderly females in th... [4]; Artificial Intelligence in Criminal Proceedings: Ensuring Legality, Validity and Fairness of Court Verdi... [5]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of AI educational assessment and built-environment evaluation. Together they illuminate construct validity: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

The neighboring evidence base brings together Post-occupancy Evaluation of a Built Environment: The Case of Konak Square (İzmir, Turkey) [6]; Validity Arguments Meet Artificial Intelligence in Innovative Educational Assessment [7]; Healing environment correlated with patients' psychological comfort: Post-occupancy evaluation of genera... [8]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of AI educational assessment and built-environment evaluation. Together they illuminate data drift: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

A complementary strand is visible in AI in essay-based assessment: Student adoption, usage, and performance [9]; Post-occupancy evaluation: A diagnostic tool to establish and sustain inclusive access in Kyrenia Town C... [10]; Student behavior in a web-based educational system: Exit intent prediction [11]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of AI educational assessment and built-environment evaluation. Together they illuminate fairness: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

For triangulation, the literature includes Post-occupancy evaluation and human-centered indoor environmental quality in higher-education buildings:... [12]; Explainable Artificial Intelligence for Student Academic Performance Prediction Using Random Forest and... [13]; A post-occupancy evaluation of a green rated and conventional on-campus residence hall [14]. These publications were retrieved and verified through DOI or publisher metadata, and their titles directly intersect the problem boundary of AI educational assessment and built-environment evaluation. Together they illuminate explainability: some emphasize representations or mechanisms, whereas others foreground validation and application conditions. The useful comparison is not a leaderboard across incompatible settings, but a structured reading of what each study controls, leaves variable, and treats as a meaningful outcome. Divergence can then reveal a change in scale, data process, endpoint, or operating constraint.

5. Operational Implications

Operationalization begins with a staged sequence. First, define the decision and its failure cost. Second, specify the evidence and operating envelope. Third, choose a method whose inductive assumptions match that evidence. Fourth, test explainability and teacher oversight under controlled perturbations. Finally, retain provenance so that an unexpected outcome can be traced to data, configuration, material batch, environmental condition, or user interaction. The workflow connects comparing AI-based student-performance prediction as a case study in educational assessment with built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space through evidence alignment and transfer boundaries connected to observable requirements.

The broader implication is that responsible progress in AI educational assessment and built-environment evaluation rests on interactions among components. Interfaces determine how uncertainty moves between measurement and inference, between algorithm and operator, or between microscopic design and system behavior. Joint work benefits from advance specification of acceptance criteria and falsifying evidence. When those agreements are explicit, efficiency goals remain subordinate to explicit validity, safety, and interpretation constraints.

6. Limitations and Future Directions

The review remains constrained by differences in vocabulary, protocol, and level of reporting. A valid DOI record authenticates bibliographic facts without independently certifying empirical conclusions. Fast-moving records create a further version-control problem. The supplied focal strings remain preserved while questionable normalization is shown only as a candidate.

Subsequent studies need to preregister comparison protocols where feasible, release machine-readable configurations, and evaluate transfer across at least one meaningful shift in operational constraints. Research should also organize error cases around mechanisms instead of counts alone. For AI educational assessment and built-environment evaluation, a valuable shared benchmark would combine baseline effectiveness, resource consumption, calibration, and disturbance response. Such a benchmark would reward designs that remain useful when evidence is incomplete or conditions change.

7. Conclusion

The body of studies addressing AI educational assessment and built-environment evaluation becomes clearer when treated as a linked evidence chain rather than a collection of isolated scores. The focal studies show how comparing AI-based student-performance prediction as a case study in educational assessment with built-environment indicators, spatial correlation, and a dual-perspective evaluation framework for everyday public space through evidence alignment and transfer boundaries; the additional literature reveals the assumptions required to interpret those contributions. Across construct validity, data drift, fairness, explainability, and teacher oversight, the most durable principle is alignment: the representation or mechanism must match the problem, metrics should fit the choice and real-world claims the evidence boundary. Such alignment improves reproducibility, transfer safety, and diagnostic value under failure.

References

1. Song, S., Zhang, X., Liang, X., & Xiao, B. (2026, February). Application of Artificial Intelligence in Educational Assessment: A Case Study of Student Performance Prediction. In Proceedings of the 2026 International Conference on Big Data and Informatization Education (pp. 617-621).

2. Chen, L., Wan, H., Hu, H., Zhang, H., & Liu, L. (2025). Correlation Analysis Between Everyday Public Space and Urban Built Environment and a Study of Its Evaluation Framework: Taking Tianjin as an Example. Buildings, 15(23), 4348. https://doi.org/10.3390/buildings15234348

3. Tang, L., & Chen, S. (2025). Educational data mining for student performance prediction in artificial intelligence environment. Molecular & Cellular Biomechanics, 22(5), 692. https://doi.org/10.62617/mcb692

4. Rieh, S. Y. (2018). Post-occupancy evaluation of urban public housing in Korea: Focus on experience of elderly females in the ageing society. Indoor and Built Environment, 29(3), 372-388. https://doi.org/10.1177/1420326x18782578

5. qizi, K. S. Z. (2026). Artificial Intelligence in Criminal Proceedings: Ensuring Legality, Validity and Fairness of Court Verdicts in Uzbekistan. International Journal of Law And Criminology, 06(04), 23-28. https://doi.org/10.37547/ijlc/volume06issue04-04

6. Malkoc, E., & Ozkan, M. B. (2010). Post-occupancy Evaluation of a Built Environment: The Case of Konak Square (İzmir, Turkey). Indoor and Built Environment, 19(4), 422-434. https://doi.org/10.1177/1420326x10365819

7. Dorsey, D. W., & Michaels, H. R. (2022). Validity Arguments Meet Artificial Intelligence in Innovative Educational Assessment. Journal of Educational Measurement, 59(3), 267-271. https://doi.org/10.1111/jedm.12331

8. Mahmood, F. J., & Tayib, A. Y. (2019). Healing environment correlated with patients' psychological comfort: Post-occupancy evaluation of general hospitals. Indoor and Built Environment, 30(2), 180-194. https://doi.org/10.1177/1420326x19888005

9. Smerdon, D. (2024). AI in essay-based assessment: Student adoption, usage, and performance. Computers and Education: Artificial Intelligence, 7, 100288. https://doi.org/10.1016/j.caeai.2024.100288

10. Güvenbaş, G., & Polay, M. (2020). Post-occupancy evaluation: A diagnostic tool to establish and sustain inclusive access in Kyrenia Town Centre. Indoor and Built Environment, 30(10), 1620-1642. https://doi.org/10.1177/1420326x20951244

11. Kassak, O., Kompan, M., & Bielikova, M. (2016). Student behavior in a web-based educational system: Exit intent prediction. Engineering Applications of Artificial Intelligence, 51, 136-149. https://doi.org/10.1016/j.engappai.2016.01.018

12. Ghiai, M., Bassaw, C., & Niknia, S. (2026). Post-occupancy evaluation and human-centered indoor environmental quality in higher-education buildings: a structured review. Frontiers in Built Environment, 12. https://doi.org/10.3389/fbuil.2026.1837203

13. Muhammad Hasanuddin, (2026). Explainable Artificial Intelligence for Student Academic Performance Prediction Using Random Forest and SHAP. Journal of Computer Science Artificial Intelligence and Communications, 3(1). https://doi.org/10.64803/jocsaic.v3i1.174

14. Bonde, M., & Ramirez, J. (2015). A post-occupancy evaluation of a green rated and conventional on-campus residence hall. International Journal of Sustainable Built Environment, 4(2), 400-408. https://doi.org/10.1016/j.ijsbe.2015.07.004